Appeal No. 2001-1031
Application No. 09/136,483

APPENDIX 2

Best Available Copy

3rd EDITION

QUANTITATIVE ANALYSIS

R. A. Day, Jr.
Emory University

A. L. Underwood
Emory University

Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Best Available Copy

Library of Congress Cataloging in Publication Data

Day, Reuben Alexander
Quantitative analysis.

Includes bibliographies.
1. Chemistry, Analytic—Quantitative.
I. Underwood, Arthur Louis joint author.
II. Title.
QD101.2.D37 1974 545 73-17387
ISBN 0-13-746537-8

© 1974 by PRENTICE-HALL, INC.,
Englewood Cliffs, New Jersey

All rights reserved. No part of this book may be reproduced in any form or by any means without permission in writing from the publisher.

10 9 8 7 6 5 4 3

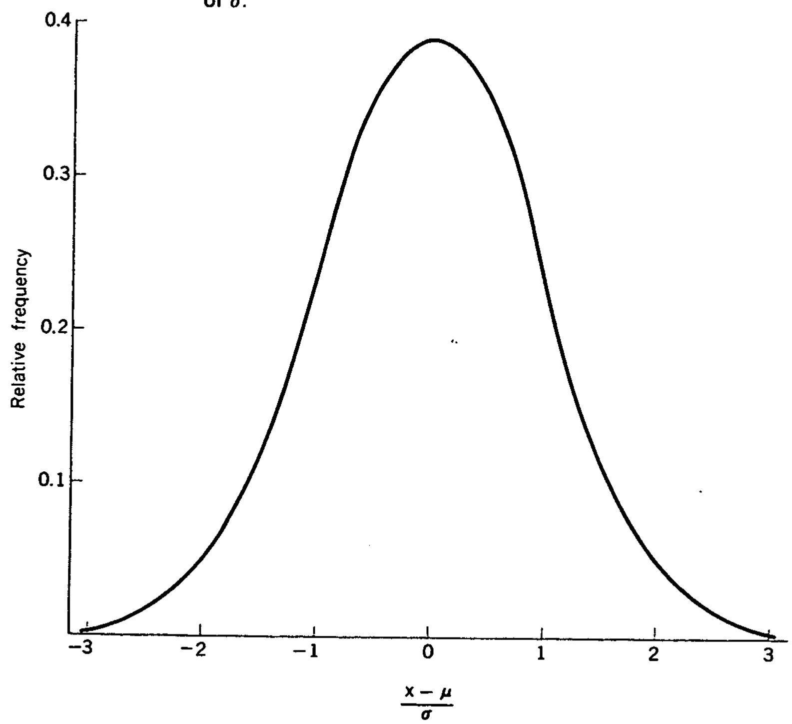
Printed in the United States of America

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

 Best Available Copy

The limiting case approached by the frequency polygon as more and more replicate measurements are performed is the *normal* or *Gaussian* distribution curve, shown in Fig. 3.2. This curve is the locus of a mathematical function which is well-known, and it is more easily handled than the less ideal and more irregular curves that are often obtained with a smaller number of observations. Data are often treated as though they were normally distributed in order to simplify their analysis, and we may look upon the normal error curve as a model which is approximated more or less closely by real data. It is supposed that there exists a "universe" of data made up of an infinite number of individual measurements, and it is actually this "infinite population" to which the normal error function pertains. A finite number of replicate measurements is considered by statisticians to be a sample drawn in a random fashion from a hypothetical infinite population; thus the sample is at least hopefully a representative one, and fluctuations in its individual values may be considered to be normally distributed, so that the terminology and techniques associated with the normal error function may be employed in their analysis.

FIGURE 3.2 Normal distribution curve; relative frequencies of deviations from the mean for a normally-distributed infinite population; deviations $(x - \mu)$ are in units of σ.



Best Available Copy

The equation of the normal error curve may be written for our purposes as follows:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Here y represents the relative frequency with which random sampling of the infinite population will bring to hand a particular value x. The quantities μ and σ, called the population parameters, specify the distribution. μ is the *mean* of the infinite population, and since we are not here concerned with determinate errors, we may consider that μ gives the correct magnitude of the measured quantity. It is clearly impractical to determine μ by actually averaging an infinite number of measured values, but we shall see below that a statement can be made from a finite series of measurements regarding the probability that μ lies within a certain interval. To the extent of our confidence in having eliminated determinate errors, such a statement approaches an assessment of the true value of the measured quantity. σ, which is called the *standard deviation*, is the distance from the mean to either of the two inflection points of the distribution curve, and may be thought of as a measure of the spread or scatter of the values making up the population; σ thus relates to precision. π has its usual significance and e is the base of the natural logarithm system. The term $(x - \mu)$ represents simply the extent to which an individual value x deviates from the mean.

The distribution function may be normalized by setting the area under the curve equal to unity, representing a total probability of one for the whole population. Since the curve approaches the abscissa asymptotically on either side of the mean, there is a small but finite probability of encountering enormous deviations from the mean. A person who happened to encounter one of these in performing a series of laboratory observations would be unfortunate indeed; some of us who have faith in never obtaining such a "wild" result in our own work are inclined to the view that the normal distribution as a model for real data breaks down, and that only the central region of the distribution curve is pertinent when applied to scientific measurements by competent workers. The area under the curve between any two values of $(x - \mu)$ gives the fraction of the total population having magnitudes between these two values. It may be shown that about two-thirds (actually 68.26%) of all the values in an infinite population fall within the limits $\mu \pm \sigma$, while $\mu \pm 2\sigma$ includes about 95% and $\mu \pm 3\sigma$ practically all (99.74%) of the values. Happily, then, small errors are more probable than large ones. Since the normal curve is symmetrical, high and low results are equally probable once determinate errors have been dismissed.

When a worker goes into the laboratory and measures something, we suppose that his result is one of an infinite population of such values that he might obtain in an eternity of such activity; then the chances are roughly 2 to 1 that his measured values will be no further than σ from the mean of the infinite population, and about 20 to 1 that his result will lie in the range

Best Available Copy

$\mu \pm 2\sigma$. In practice, of course, we can never find σ for an infinite population, but the standard deviation of a finite number of observations may be taken as an estimate of σ. Thus we may predict something about the likelihood of occurrence of an error of a certain magnitude in the work of a particular individual once he has performed enough measurements to permit estimation of the characteristics of his particular infinite population.

STATISTICAL TREATMENT OF FINITE SAMPLES

Although there is no doubt as to its mathematical meaning, the normal distribution of an infinite population is a fiction so far as real laboratory work is concerned. We must now turn our attention to techniques for handling scientific data as we obtain them in practice.

Measures of Central Tendency and Variability

The *central tendency* of a group of results is simply that value about which the individual results tend to "cluster." For an infinite population, it is μ, the mean of such a sample. The *mean* of a finite number of measurements, $x_1, x_2, x_3, \ldots, x_n$, is often designated $\bar{x}$ to distinguish it from μ. Of course $\bar{x}$ approaches μ as a limit when n, the number of measured values, approaches infinity. Calculation of the mean involves simply averaging the individual results:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{i=n} x_i}{n}$$

The mean is generally the most useful measure of central tendency. It may be shown that the mean of n results is $\sqrt{n}$ times as reliable as any one of the individual results. Thus there is a diminishing return from accumulating more and more replicate measurements: The mean of four results is twice as reliable as one result in measuring central tendency; the mean of nine results is three times as reliable; the mean of twenty-five results, five times as reliable, etc. Thus, generally speaking, it is inefficient for a careful worker who gets good precision to repeat a measurement more than a few times. Of course the need for increased reliability, and the price to be paid for it, must be decided on the basis of the importance of the results and the use to which they are to be put.

The *median* of an odd number of results is simply the middle value when the results are listed in order; for an even number of results, the median is the average of the two middle ones. In a truly symmetrical distribution, the mean and the median are identical. Generally speaking, the median is a less efficient measure of central tendency than is the mean, but in certain instances it may be useful, particularly in dealing with very small samples.

Since two parameters, μ and σ, are required to specify a frequency distribution, it is clear that two populations may have the same central tendency